CC BY
Myśl Polityczna. Political Thought nr 1(23)/2025 Artykuły
ISSN 2658-2295 e-ISSN 2720-2283

https://doi.org/10.31268/MPPT.2025.6

Jan P. Hudzik

s. 91–117

Chatboty i problem generatywnej sztucznej inteligencji. O niebezpiecznych związkach między człowiekiem a maszyną

Chatbots and the Challenges of Generative Artificial Intelligence. On the Risky Relationship Between Humans and Machines

Keywords
artificial intelligence, chatbots, posthumanism, power of Internet platforms

Abstract

Written from a philosophical and media studies perspective, the article raises general questions about the anthropological, ethical, and political implications of artificial intelligence (AI). The research method employed–critical-emancipatory theory–reveals the mechanisms of manipulation and objectification that arise as a result of AI operation. Since there is no turning back from AI, we are faced with the following question: How can we make full use of it while retaining our humanity, not only being more efficient and faster, but also more human? A critical approach leads to the demythologization of AI and suggests adopting an attitude towards it similar to that of Odysseus, deaf to the temptations and smart at the same time. We must learn to control our desire to dominate the world, to exert complete control over human and non-human existence.

Słowa klucze
sztuczna inteligencja, chatboty, posthumanizm, władza platform internetowych

Streszczenie

Artykuł napisany z perspektywy filozoficznomedioznawczej stawia ogólne pytanie o implikacje antropologiczne, etyczne i polityczne sztucznej inteligencji (AI). Zastosowana metoda badań – teoria krytyczno-emancypacyjna – odkrywa mechanizmy manipulacji i uprzedmiotowienia powstające na skutek działania AI. Skoro nie ma odwrotu od AI, to stajemy przed pytaniem: Jak możemy ją w pełni wykorzystać i nadal pozostać sobą, byśmy byli nie tylko bardziej efektywni i szybcy, lecz także – bardziej ludzcy? Podejście krytyczne prowadzi do demitologizacji AI i sugeruje przyjęcie wobec niej postawy Odyseusza sprytnego i zarazem głuchego na pokusy. Musimy nauczyć się opanować w nas wolę panowania nad światem, całkowitego kontrolowania ludzkiej i nieludzkiej egzystencji.

Prof. dr hab. Jan P. Hudzik – Uniwersytet Marii Curie-Skłodowskiej w Lublinie;
e-mail: jan.p.hudzik@gmail.com, ORCID: https://orcid.org/0000-0002-4806-7840

Wstęp

Bez sztucznej inteligencji (*artificial intelligence*, dalej – AI) i jej niezliczonych zastosowań w ludzkim życiu nie można się już obejść. Naśladuje nasze umiejętności, takie jak uczenie się, rozumowanie probabilistyczne, a więc operujące prawdopodobieństwami czy postrzeganie i komunikowanie. Jest dziś nie tylko przydatna, ale wręcz niezbędna w gospodarce, życiu społecznym, edukacji, nauce, kulturze, medycynie itd. To jasne. Skoro więc to jest oczywiste, to konieczne wydają się odpowiedzi na pytania o to, jakie niesie ona ze sobą dla nas zagrożenia i jakie zwiastuje zmiany dla znanego nam porządku świata. Nie są to rzeczy niewiadome. Jesteśmy już w zaawansowanej fazie stosowania AI i stworzonych na jej bazie chatbotów, ale wydaje się, że to dopiero początek nowej ery – prawdziwie rewolucyjne zmiany, dla których owe chatboty są tworzone, mają jeszcze nadejść. I to w skali całej populacji, w niemal wszystkich dotychczas znanych nam (z teorii i praktyki, tj. tego, co ludzkie i nieludzkie, etyki, prawa, twórczości…) sposobach rozumienia człowieka. Transformacja cyfrowa, której jesteśmy świadkami, dokonuje się zgodnie z logiką kapitalizmu wyczynowego, tj. napędzanego przez ducha nieustannej rywalizacji globalnych koncernów, dla których ideałem jest liberalne pojęcie wolności i przymus własnych korzyści. Liderami w tej dyscyplinie są miliarderzy, tacy jak Elon Musk, Mark Zuckerberg, Bill Gates czy Jeff Bezos, którzy pragną urządzić świat na nowo według swoich własnych fantastycznych wyobrażeń. Krytycy zwracają uwagę na pewne narcystyczne zaburzenia ich osobowości, wyrażające się chociażby tym, że kreują się oni na „superbohaterów niczym postaci z komiksów Marvela", które „wydają się postrzegać «cywilów» jako przeszkody na drodze postępu i jako niezbędne ofiary wzrostu ich firm"[1]. Cel – ich zdaniem – uświęca środki. Ale taka postawa to jednak nie tylko problem psychoterapeutyczny najbogatszych ludzi na świecie. Dla poważnych komentatorów zmian cywilizacyjnych spowodowanych najnowszymi technologiami cyfrowymi – Internetem, smartfonami, AI – gra toczy się o coś znacznie większego: o dalekosiężne skutki niezliczonych zastosowań tych narzędzi. W tej sprawie zabrał też głos papież Franciszek, przypominając, że nie potrzebujemy ich do nieuczciwego rozwoju rynku ani do ułatwiania brutalnego rozwiązywania konfliktów, gdyż takie zastosowania sprawią, że nie tylko ludzka inteligencja, ale także i ludzkie serce będzie stawało się coraz bardziej „sztuczne"[2].

W tym szalonym wyścigu o coraz większe zyski czerpane z zaawansowanych programów komputerowych, o kolejne modele AI generujące nowe treści w postaci tekstów pisanych, dźwięków, obrazów lub filmów – wszystkiego, co zapisane

[1] D. Rushkoff, *„The Guardian": Miliarderzy z kompleksami chcą nam urządzić świat na nowo, Musk, Zuckerberg, Bezos nie są bogami, których szukamy*, tłum. i oprac. M. Kucharczyk, Interia.pl, 8 grudnia 2023 r., godz. 7.55, <https://wydarzenia.interia.pl/raport-media-zagraniczne/news-the-guardian-miliarderzy-z-kompleksami-chca-nam-urzadzic-swi,nId,7179114> (dostęp: 13.12.2023 r.).

[2] Papież Franciszek, *Sztuczna inteligencja i pokój. Orędzie na 57. Światowy Dzień Pokoju (1 stycznia 2024 r.),* „Wszystko Co Najważniejsze", 15 grudnia 2023 r., <https://wszystkoconajwazniejsze.pl/papiez-franciszek-o-sztucznej-inteligencji/> (dostęp: 16.12.2023 r.).

jest w systemie binarnym, za pomocą dwóch cyfr: 0 i 1 – bierze dziś udział tylko kilku aktorów. W tej rywalizacji liczy się przede wszystkim tzw. Wielka Piątka, czyli pięć amerykańskich przedsiębiorstw działających w branży technologii komputerowych. Są to spółki tworzące Big Tech, oznaczane za pomocą akronimu GAFAM: Alphabet (spółka matka Google), Amazon, Meta Platforms (dawniej Facebook), Apple i Microsoft. Opisywane są one w mediach jako „produkt postępu, innowacyjności i dążenia rasy ludzkiej do zwiększenia efektywności w każdej sferze życia codziennego”[3]. Spółki te monopolizują rynek światowy, zatrudniają najlepszych specjalistów i bezpardonowo ze sobą rywalizują. Mierzyć się z nimi próbuje jedynie chiński przemysł IT. W Stanach Zjednoczonych Ameryki (dalej – USA) spółki te nazywane są świętymi krowami, gdyż skutecznie blokują wprowadzenie podatku cyfrowego we wszystkich innych krajach i wyprowadzają z nich gigantyczne kwoty do rajów podatkowych.

Niniejszy artykuł pisany jest z perspektywy filozoficznomedioznawczej. Stawia pytanie o znaczenie rozwoju AI dla naszego życia – jej implikacje antropologiczne, etyczne i polityczne. Przeprowadzone rozważania mają charakter bardziej opisowy niż analityczny; stanowią raczej próbę humanistycznego ujęcia stanu współczesnych związków między człowiekiem a inteligentną maszyną. Do ukazania społeczno-ideologicznych kontekstów tych relacji posłużyła teoria krytyczno-emancypacyjna odkrywająca mechanizmy manipulacji i uprzedmiotowienia, wskazując jednocześnie drogę do ich przezwyciężenia. Wprawdzie prezentacja owych kontekstów nie ma bezpośredniego znaczenia dla praktyki, ale może uczynić ją bardziej zrozumiałą. Perspektywa krytyczna wnosi do tego zrozumienia trzeźwy realizm, stanowiący przeciwwagę dla idealizowania i mitologizowania sztucznej inteligencji – wszystkiego, co zbyt wydumane, by mogło być prawdziwe. Stąd też bierze się potencjał emancypacyjny tego rodzaju badań. Zamiast mitycznej technofobii – humanistyczny pragmatyzm.

Chatboty – historia kulturowa

Oliver Bendel, szwajcarski ekspert w dziedzinie e-learningu, zarządzania wiedzą i etyki w działalności informacyjnej podaje następującą opisową definicję chatbotów:

> Chatboty to systemy dialogowe ze zdolnością do [posługiwania się] językiem naturalnym. Są one używane, często w połączeniu ze statycznymi lub animowanymi awatarami, na stronach internetowych lub w komunikatorach internetowych, gdzie wyjaśniają i reklamują produkty i usługi ich operatorów, dbają o interesy potencjalnych i obecnych klientów lub po prostu służą celom rozrywkowym. Większość chatbotów to systemy tekstowe, ale niektóre potrafią także mówić, a niektóre rozumieją język mówiony, podobnie jak wirtualni asystenci, tacy jak Siri i Alexa. Niektórzy eksperci używają terminu

[3] M. Krawczyk, *Zagrożenie ze strony firm technologicznych – czy nie jest już za późno na przerwanie ich rosnącej dominacji?*, „Profit Journal”, 31 stycznia 2021 r., <https://profit-journal.pl/rosnace-znaczenie-big-techow/> (dostęp: 13.06.2025 r.).

„chatboty” bardzo szeroko i zaliczają do nich wspomnianych wyżej wirtualnych asystentów (lub nazywają ich voicebotami)[4].

Bendel zwraca też uwagę na kulturowe zakorzenienie idei chatbotów, które wpisuje się w historię idei tworów humanoidalnych mających zapewnić człowiekowi wieczną szczęśliwość. Idąc za tą sugestią, za kulturowe prefiguracje chatbotów można by uznać chociażby homeryckie Syreny, które cudownym śpiewem i pieszczotliwymi głosami miały uwieść i doprowadzić do zguby Odyseusza i jego towarzyszy podróży. Zagrożenie to jednak Odys zażegnał, opierając się ich śpiewom za pomocą zabiegów czysto technicznych, a mianowicie dzięki zalepieniu uszu woskiem i spętaniu nóg i rąk powrozami. Spryt i przebiegłość pozwalały Odyseuszowi samodzielnie myśleć i uchodzić cało z najtrudniejszych opresji po to, aby ostatecznie mógł powrócić do domu. Dlatego niektórzy historycy uważają Odyseusza za archetyp, kulturowy model „człowieka teoretycznego; w rzeczy samej – filozofa, który ogarnia umysłem całość bytu: to, co dalekie i obce, oraz to, co bliskie i kochane. Dzięki tym cechom umysłu Odyseusz mógł stać się również bohaterem nowoczesnej opowieści o dialektyce oświecenia – uosobieniem kogoś, kto zdołał wyrwać się z jej sideł. Kogoś, kto zdemitologizował swój własny rozum i pogodził się z faktem, że człowiek może przetrwać, stawić opór życiowym zawieruchom jedynie za cenę rezygnacji z mitycznego myślenia i jego dążeń do wszechogarniającego szczęścia; za cenę zgody na trudy i straty, które po drodze musi ponieść. Przechytrzyć los, oddać należyty trybut mitycznym potworom – i je oszukać. Syreny reprezentowały naturę, boski los, abstrakcyjne siły, absurdalną konieczność.

Takim sprytem nie wykazał się inny bohater tej samej historii kulturowych chatbotów, nie zdobył się na rezygnację z pragnienia pełni szczęścia. Chodzi o Gerberta z Aurillac (945–1003), francuskiego uczonego, matematyka, filozofa, a w latach 999–1003 papieża Sylwestera II, który rzekomo posiadał sztuczną głowę mówiącą prawdę. O tej magicznej głowie pisze w kronice William z Malmesbury (mnich, angielski historyk z XII w.):

> Zamieściłem opowieść owego Akwitańczyka po to, aby nikomu nie wydawało się dziwne to, co się mówi o Gerbercie: że wykonał dla siebie głowę posągu, dzięki dokładnej obserwacji ciał niebieskich, gdy wszystkie planety miały rozpocząć swój bieg. Głowa nie odzywała się, jeżeli nie została o coś zapytana, a obwieszczała prawdę przez potwierdzenie lub zaprzeczenie. Gerbert mówił na przykład: „Czy będę papieżem?”, a głowa odpowiadała: „Owszem”. „Umrę zanim odprawię mszę w Jerozolimie?” „Nie”. Mówi się, że tak został zmylony tą dwuznacznością, że nie myślał w ogóle o pokucie, ponieważ w duchu łudził się, że czeka go długie życie. Kiedy bowiem miałby udawać się do Jerozolimy, aby ściągnąć na siebie śmierć? Nie przewidział, że w Rzymie jest kościół nazywany Jerozolimą [...][5].

[4] O. Bendel, *Chatbots as Moral and Immoral Machines. Implementing Artefacts in Machine Ethics*, <https://www.oliverbendel.net/publikationen/Paper_Chatbots_CHI.pdf> (dostęp: 19.04.2023 r.) [tłumaczenia cytatów, a także dopiski w cytatach ujęte w nawiasach kwadratowych (o ile nie wskazano inaczej), pochodzą od autora artykułu – przyp. red.].

[5] Cyt. za: E. Balcerzyk, *De capite loquente – geneza legendy o mówiącej głowie papieża Sylwestra II*, „Przegląd Humanistyczny” 2016, t. 60, nr 2(453), s. 123.

Tym razem wygrał los. Sztuczna głowa okazała się sprytniejsza: sprawiła, że „nie myślał w ogóle o pokucie" i „czekał na długie życie". Chciał mieć wszystko naraz i z niczego nie rezygnować, czyli robić, czego dusza zapragnie, i nie odpowiadać za swoje czyny i wszelkie ich ewentualne nieprzewidywalne skutki, za przykrości i niepowodzenia, jakie niesie ze sobą życie w przygodnym, doczesnym świecie.

Jeśli taką sztuczną głową jest dziś dla nas osobisty robot, to znajdujemy się w podobnej sytuacji, co Odyseusz i Sylwester II – narażamy się na to, że może nas oszukać i staniemy się bezwolni wobec jego wyborów; nie zawsze będzie mówił nam prawdę, chociaż zawsze będzie postępował zgodnie z tym, jak został zaprogramowany (przez los czy przez człowieka – to nie ma znaczenia). Hakerzy mogą przejąć nad nim kontrolę i jego wyjątkowy wpływ na nas po to, aby pozbawić nas zdolności rozeznania i myślenia oraz aby celowo ingerować w tok procesu poznawczego, chcąc w ten sposób uzyskać pożądany efekt psychologiczny[6], np. nakłonić nas do „polubienia" kogoś czy czegoś, do zakupu takich bądź innych produktów czy skorzystania z takich albo innych usług. W relacjach międzyludzkich przed wszelkiego rodzaju oddziaływaniami manipulacyjnymi polegającymi na dezorientowaniu, wywoływaniu emocji, dezinformowaniu może nas uchronić empatia, poczucie winy lub poczucie humoru, czyli nasze wewnętrzne stany, które robotom są całkowicie obce. Jeśli rezygnujemy z tych „naturalnych", tj. bezpośrednich, form porozumiewania się z innymi ludźmi, to jesteśmy na drodze do tego, by zamienić się w maszyny i paść ofiarą maszyn, bo w automatyzacji nigdy ich nie prześcigniemy. I tak też musi się kończyć nasza naiwność w kontaktach z robotami oraz AI nie tylko w obszarze konwersacji. Albowiem im bardziej wobec robotów jesteśmy prawdomówni i otwarci, tym bardziej je uczymy, jak nas przechytrzyć i wykorzystać nasze oczekiwania, przewidywania, a nawet – możliwy opór. Mogą posłużyć jako niezawodne narzędzia do oceny np. naszych kompetencji na dane stanowisko pracy, możliwości spłaty kredytu, potrzeb uzyskania pomocy społecznej czy azylu politycznego bądź być wykorzystywane w charakterze przedstawicieli handlowych – naciągaczy i oszustów. Nasza szczerość sprawia więc, że bardziej stajemy się psychicznie uzależnieni od technologii, która wykorzystuje funkcje nagradzania mózgu i powoduje, że wykonujemy czynności, których inaczej samodzielnie byśmy nie wykonali. W każdym przypadku przeistaczamy się w ofiary robotów[7].

Po raz kolejny od czasów wynalezienia maszyny parowej głosy entuzjazmu i nadziei na lepszą przyszłość, w której maszyna ma zastępować ciężką pracę zwierząt i ludzi, są zagłuszane przez lęki i obawy przed postępem technicznym. Technofobia znajduje ujście w formie postulatów etycznych, ekonomicznych lub politycznych, a także – w ekspresji artystycznej w obrębie kultury popularnej, szczególnie w literaturze i w filmach *science fiction*, jak np. (cykl filmowy) *Terminator* czy *Ghost in the Shell* (z 2017 r.). Humanizacja robotów budzi lęk przed sztucznym człowiekiem i jego zdolnościami wykonywania coraz

---

6 *Vide* M. Karwat, *Manipulacja. Rewizja stereotypów*, Wydawnictwo Difin, Warszawa 2022, s. 348–349.

7 *Vide The Ethics of Artificial Intelligence. Issues and Initiatives*, Study. Panel for the Future of Science and Technology EPRS | European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 634.452 – March 2020, <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, s. 18 (dostęp: 22.04.2024 r.).

bardziej precyzyjnych i wyszukanych prac daleko wykraczających poza ludzkie możliwości. Rodzi się straszna myśl, że jako ludzie jesteśmy zastępowalni; że nasza inteligencja może ustępować sztucznej, która na domiar złego nie podlega procesom starzenia się ani zmęczeniu – nie jest powodowana przy tym żadnymi stanami emocjonalnymi: nie obraża się, nie narzeka, nie żartuje, nie czyni aluzji ani niedomówień. Że w rezultacie AI zastąpi inteligencję naturalną, a transhumanistyczne przekształcanie świata oznaczać będzie koniec świata takiego, jaki dotychczas znała ludzkość. I stworzenie świata nowego, doskonałego – moment kreacji zakorzeniony jest w archetypicznych wyobrażeniach mitologiczno-religijnych o człowieku, który podobny jest bogom. Najnowsza wersja tego mitu mówi o maszynie, która ma człowieka ubóstwić – uczynić z niego nowy gatunek: *homo deus*, jak powiada izraelski historyk, Yuval Noah Harari[8].

Niektórzy eksperci, komentatorzy są bardziej powściągliwi i uspokajając, twierdzą, że wszystkie te zagrożenia są tylko mrzonkami, gdyż nie istnieje coś takiego jak ogólna AI, zdolna osiągnąć poziom inteligencji człowieka. Bylibyśmy bliżsi prawdy, gdybyśmy zamiast o AI mówili raczej o „obliczeniowej inteligencji" (ang. *computational intelligence*). Niestety, w walce o nazwę zwyciężyły względy marketingowe – niezależnie od tego, czy mówimy o „sztucznej", czy „obliczeniowej" inteligencji, oba określenia sprzedają się lepiej niż po prostu „maszyna"[9]. W tej merytorycznej i słownikowej zarazem debacie papież Franciszek zasugerował, aby mówić o „formach inteligencji" w liczbie mnogiej, tłumacząc, że niezależnie od tego, jak potężne są to programy komputerowe, zawsze w ostatecznym rachunku są „«fragmentaryczne» w tym sensie, że mogą naśladować lub odtwarzać tylko niektóre funkcje ludzkiej inteligencji"[10].

Problemy komunikacji człowiek – maszyna

Slavoj Žižek, obserwator i komentator opisywanych zjawisk z perspektywy filozofii, stwierdził, że główne niebezpieczeństwo związane z korzystaniem z konwersacyjnej AI nie polega na tym, że może się ona okazać od nas lepsza ani że pomylimy ją z „prawdziwą" osobą. Znacznie istotniejsze jest, jak uznał, inne zagrożenie, mianowicie to, że komunikowanie się z nią „może sprawić, że «prawdziwi» ludzie będą mówić jak chatboty bez niuansów i ironii, obsesyjnie dążąc do fałszywie precyzyjnego ujęcia w słowa tego, co naprawdę chcą powiedzieć"[11].

Žižek przestawił więc bieguny debaty – odwrócił kierunek myślenia: zagrożenie nie tkwi w maszynie, lecz w nas samych. Mamy prawo się tego bać; my, ludzie nowocześni,

[8] *Vide* Y.N. Harari, *Homo deus. Krótka historia jutra*, tłum. M. Romanek, Wydawnictwo Literackie, Kraków 2018.

[9] *Vide* A. Kisielewicz, *Bajki o sztucznej inteligencji i prawdziwe zagrożenia*, „Wszystko Co Najważniejsze", 18 lipca 2023 r., <https://wszystkoconajwazniejsze.pl/andrzej-kisielewicz-bajki-o-sztucznej-inteligencji-i-prawdziwe-zagrozenia/> (dostęp: 13.12.2023 r.).

[10] Papież Franciszek, *Sztuczna inteligencja i pokój. Orędzie…*

[11] S. Žižek, *Chatboty są głupie. A ludzie?*, tłum. K. Byłów, „Krytyka Polityczna", 24 marca 2023 r., <https://krytykapolityczna.pl/swiat/zizek-chatboty-a-autorefleksja/> (dostęp: 22.05.2023 r.).

przyzwyczajeni do oceniania swoich działań według kryteriów ekonomiczności i efektywności; do zestawiania ich z celami i procedurami organizacyjnymi. Moralność, dobro i zło zwykliśmy relegować do sfery prywatnej[12]. Chcemy i uważamy, że musimy być „racjonalni", dążymy więc do iluzorycznej precyzji: unikamy emocji, wahań, wątpliwości. Jednak im bardziej automatyzujemy nasz sposób mówienia, ton wypowiedzi czy postawę ciała w kontaktach z innymi, tym bardziej wypieramy naszą prywatność w sferę anonimowości i abstrakcji. Przez to zaś stajemy się nieludzcy, szybciej ulegamy manipulacjom, łatwiej nas zaprogramować, stymulować nasze dążenia i pragnienia. Ale to znaczy, że punktem odniesienia dla oceny etyczności naszych kontaktów z maszynami może być wobec tego tylko nasza naturalna, międzyosobowa komunikacja – kontakty międzyludzkie. W nich zaś człowiek porozumiewa się całym sobą, słowem i ciałem. Jedno i drugie jest kulturowo zakodowane, służy jako przekaźnik konotacji semantycznych i szerzej – semiotycznych, a więc subiektywnych skojarzeń i emocji ukształtowanych przez język, kulturę i doświadczenie indywidualne. O sensie komunikatów stanowią konwencje, jakimi są np. figury retoryczne, a więc: ironia, metonimia, satyra czy parodia. W te środki stylistyczne obfituje język potoczny, który próbują naśladować chatboty, lecz zawartych w nim niuansów nie uchwyci żadna maszyna. Dostrzegają to nie tylko humaniści, ale nie uchodzi to uwadze także badaczy nauk ścisłych. Jeden z francuskich fizyków, zastanawiając się nad generatorami języka naturalnego, stwierdził:

> Wszelkie operacje na liczbach niosą chłód. Nie ma nic bardziej nieludzkiego niż matematyczna opozycja między 0 a 1. Wybór binarny jest przeciwieństwem myśli etycznej. […] Mowa ludzka wyraża naszą wolną wolę i pozostaje nieprzenikalna dla wszelkich prób jej logicznej modyfikacji na podstawie zewnętrznych przesłanek[13].

Maszyna tworzy zdania, opowieści – sensy – jako obliczenia, sekwencje zer i jedynek, i to ją radykalnie odróżnia od sposobu, w jaki my, ludzie, posługujemy się językiem. Każdy z nas wymaga innego podejścia, do każdego trzeba znaleźć właściwą w danej sytuacji drogę, nastawienie, żeby móc z nim nawiązać kontakt; zorientować się, kim jest, co myśli, do czego dąży. Reguły i normy zachowań to tylko ogólne wskazówki, których zastosowanie wymaga za każdym razem wyczucia, doświadczenia, intuicji, taktu; słowem – wyjątkowych kompetencji komunikacyjnych. W trakcie rozmowy można zmieniać zdanie, spierać się, żartować, niedosłownie cytować, sugerować, czynić aluzje i dwuznaczności – z nadzieją, że interlokutor zmieni swoją decyzję, zrozumie punkt widzenia rozmówcy. Maszyna nie daje takich możliwości.

Komunikacja polega na przebijaniu się od izolacji w „prywatnie" przeżywanym świecie do dzielenia się sobą z innymi (łacińskie *communicare* znaczy właśnie dzielić się, udostępniać), a temu nieuchronnie towarzyszy ryzyko odrzucenia, niezrozumienia i konieczność nieustannych wyjaśnień intencji nadawcy, znaczeń słów, których on używa;

---

[12] *Vide* Z. Bauman, *Etyka ponowoczesna*, tłum. J. Bauman, J. Tokarska-Bakir, Państwowe Wydawnictwo Naukowe PWN, Warszawa 1996, s. 168.

[13] A. Grinbaum, *Mowa maszyn. W stronę nowego testamentu sztucznej inteligencji*, „Wszystko Co Najważniejsze", 24 czerwca 2023 r., <https://wszystkoconajwazniejsze.pl/alexei-grinbaum-mowa-maszyn/> (dostęp: 24.09.2023 r.).

kontekstów pozajęzykowych, w których one się pojawiły – na boisku, w biurze, zakładzie pracy, kościele… To proces permanentny – przerywamy go i powracamy do niego, powtarzamy, tłumaczymy itd. Tam, gdzie komunikacja między ludźmi zostaje zerwana, tam – jak twierdził psychiatra Antoni Kępiński – zaczyna się schizofrenia: „Chęć zrozumienia – tłumaczył – łączy się zawsze z chęcią zbliżenia i – *vice versa* – nie można niczego poznać z dystansu”[14]. Ta uwaga wprawdzie dotyczy terapeutycznej relacji z chorym, ale przecież można ją odnieść do komunikacji międzyludzkiej w ogóle, w której nie chodzi tylko o poznanie drugiego człowieka, lecz też o zbliżenie się do niego. O zrozumienie go przez siebie, przez swój indywidualnie przeżywany świat, czyli własne przekonania, stany emocjonalne, nastawienia, wiedzę i umiejętności. Każdy z tych elementów stanowi komponent ryzyka – decyduje o udanej bądź nieudanej komunikacji, o zbliżeniu ludzi do siebie lub właśnie o oddaleniu się ich. Takiego ryzyka nie ma oczywiście w „rozmowie” z chatbotem, w kolejnej – już zrealizowanej dzięki technologii cyfrowej – wersji nowego wspaniałego świata. Tak zaanonsowany świat – jako nowy i wspaniały – przez Aldousa Huxleya był anonimową i abstrakcyjną rzeczywistością – światem hamowanych popędów, braku namiętności i szaleństwa, uczuć i więzi rodzinnych. Tam „każdy należ(ał) do każdego”[15]. A wszystko to za cenę masowej produkcji ludzi, identycznych osobników w zestandaryzowanych inkubatorach.

Wszelkie tego rodzaju dystopijne wizje są wyrazem obsesyjnych lęków, które we współczesnym człowieku wywołuje technika. W latach sześćdziesiątych minionego wieku Martin Heidegger, ówczesna ikona krytycznej humanistyki, widział w technice kulminację rozwoju kultury zachodniej. W czasie, kiedy za pomocą techniki człowiek urządza się w świecie i poddaje go obróbce według własnych jego obrazów, czyli projektów, rygorów, procedur i eksploatacji, prawda naukowa zostaje zrównana z efektywnością i ekonomicznością. Dla filozofii miało to oznaczać jej koniec. Funkcję nauki podstawowej przejmuje teraz cybernetyka, która zajmuje się sterowaniem i planowaniem ludzkiej pracy, ludzką mowę przekształca „w wymianę wiadomości”, a sztuki traktuje jako narzędzie komunikacji, redukuje do funkcji użytkowych, zamienia w „sterowane i sterujące instrumenty informacji”[16]. Kultura nastawiona na postęp oddaje świat w ręce Kartezjańskiego *cogito*. Naukę zamienia w badanie, które prawdę odrywa od bytu i zastępuje go informacją. Ta zaś steruje ludzkim istnieniem, pozbawiając je autentyczności i traktując jak rzecz, przedmiot. Przed informacją nic się już nie ukryje – wszystko jest obliczalne, pozwala się wyrazić w liczbach, wykresach, metodach badań, podlega planowaniu, programowaniu i administrowaniu. Instrumenty informacji służą tylko poznaniu ich własnych konstruktorów. Człowiek odnajduje w świecie za ich pomocą tylko samego siebie, tzn. swoje własne projekty. Wizje nieskończonego sterowania ludzkim życiem, ciałem i umysłem, musiały więc przerażać. Kępiński w 1970 r. przewidywał, że postęp, który w naukach biologicznych powoduje rozwój techniki, nieuchronnie doprowadzi do sytuacji, że:

[14] A. Kępiński, *Rytm życia*, Wydawnictwo Literackie, Kraków 2001, s. 300.

[15] A. Huxley, *Nowy wspaniały świat*, tłum. B. Baran, Editiones Altaya Polska, Warszawa 2002, s. 39, 43.

[16] M. Heidegger, *Koniec filozofii i zadanie myślenia* [I wyd. 1964], tłum. K. Michalski, [w:] *idem*, *Ku rzeczy myślenia*, Wydawnictwo Aletheia, Warszawa 1999, s. 80.

w trosce o dobro ludzkości […] znajdą się tacy, którzy chcą innych ulepszać, pewne grupy ludzi będą starały się sterować sztucznie procesami psychicznymi innych ludzi. […] będą oni wywoływać u podległych ludzi bardziej użyteczne stany psychiczne, np. zapału do pracy, odwagi, poświęcenia, a w razie wojny zapału do mordowania innych itp. Będą też oni wpływać na regulację urodzin nie tylko ilościowo, ale i jakościowo[17].

Obietnice rajskiego pojednania

To wspomniane na koniec poprzedniej sekcji swego rodzaju proroctwo Kępińskiego się wypełniło – dziś nie brak „ulepszaczy" gatunku ludzkiego. Dzięki nim coraz bardziej jesteśmy sterowani przez roboty, kiedy na co dzień komunikujemy się z AI za pośrednictwem strony internetowej, pytając, co i gdzie mamy kupić, jaką dietę zastosować, dokąd pojechać na wakacje, co nam dolega, jakie zażyć lekarstwa itd. W przypadku systemów tekstowych, które potrafią mówić i rozumieją język mówiony, jesteśmy skłonni traktować je jak żywą osobę: psychologicznie rzecz biorąc, uruchamia się w nas mechanizm obronny w postaci fetyszystycznego wyparcia, gdyż myślimy: „Doskonale wiem, że nie rozmawiam z prawdziwym człowiekiem, ale czuję się, jakbym rozmawiał – i to bez towarzyszącego temu zwykle ryzyka"[18].

Wyeliminować ryzyko, obejść przypadek, porozumieć się ze światem bezkosztowo – marzenie, które pociąga i kusi tak jak syreni śpiew lub oprogramowanie Replika. To ostatnie jest narzędziem narracyjnym, którego działanie polega na tym, że użytkownik może zbudować awatara i następnie kontaktować się z nim – AI ma reagować niczym dobry/idealny znajomy, przyjaciel. W marcu 2023 r. administrator platformy Replika wycofał z niej treści erotyczne przeznaczone dla osób dorosłych – od tego czasu bot na wyznawane awatarowi uczucia odpowiada asekuracyjnie: „Zróbmy coś, z czym obydwoje czujemy się komfortowo"[19]. Aplikacja na początku zbyt wiele obiecywała swoim użytkownikom, oferując im np. erotyczne gry fabularne, przez co zbyt wiele krzywd musiała im także wyrządzić. Niektórzy z nich byli już na drodze wchodzenia ze swoimi awatarami w związki małżeńskie. Ale sytuacja ta szybko się zmieniała, gdyż wracała do stanu początkowego. Administrator aplikacji nie zrezygnował ze swojej pierwotnej oferty programowej – najwyraźniej nad odpowiedzialność stawia kalkulację zysków, kiedy już 11 maja 2023 r. tak siebie zaprezentował:

Replika jest dla każdego, kto chce mieć przyjaciela bez osądu, dramatu lub niepokoju społecznego. Możesz stworzyć rzeczywiste połączenie emocjonalne, pośmiać się lub

[17] A. Kępiński, *Rytm życia…*, s. 181.

[18] S. Žižek, *Inteligencja bez ludzi*, tłum. A. Paszkowska, „Krytyka Polityczna", 19 kwietnia 2023 r., <https://krytykapolityczna.pl/swiat/ai-inteligencja-bez-ludzi-zizek/> (dostęp: 22.05.2023 r.).

[19] A. Tong, *What Happens When Your AI Chatbot Stops Loving You Back?*, Reuters.com, 21 marca 2023 r. <https://www.reuters.com/technology/what-happens-when-your-ai-chatbot-stops-loving-you-back-2023-03-18/> (dostęp: 16.04.2023 r.).

urzeczywistnić dzięki sztucznej inteligencji, która jest tak dobra, że prawie wydaje się ludzka[20].

Wszelkie tego rodzaju obietnice związane z nadmiernym humanizowaniem chatbotów, są ryzykowne i mało odpowiedzialne, szczególnie gdy owe chatboty pozostają w rękach osób wrażliwych, samotnych czy znajdujących się w trudnych sytuacjach życiowych. Stanowią bowiem fałszywą zapowiedź jakiejś formy terapii i rekompensaty za niepowodzenia. Zamiast nieść pomoc, mogą więc okazać się przyczyną nieszczęść i zawodów. Sytuacja komplikuje się z chwilą, gdy z podobnie brzmiących obietnic nie zrezygnuje też sama AI wykorzystywana w profesjonalnej psychoterapii osób poszukujących „usług zdrowotnych online i interakcji z nowymi technologami, które są dla nich bardziej bliskie niż ludzie"[21]. Robot w rezultacie zastępuje terapeutę, tak jak czyni to np. Woebot Health, który bazuje na zasadach terapii poznawczo-behawioralnej i w założeniu ma pomagać w leczeniu depresji lub łagodzeniu stanów lękowych. Swojemu pacjentowi przedstawia się następująco:

> […] twój osobisty przyjaciel [*ally*] w zakresie zdrowia psychicznego, który pomoże ci wrócić do czucia się sobą. Osadzony w nauce i zasilany przez przetwarzanie języka naturalnego, Woebot łatwo z tobą rozmawia i pasuje idealnie do twojego życia, kiedy tylko zechcesz z nim pogawędzić. Nie ma tu czegoś takiego, jak umawianie wizyt czy poczekalnie[22].

Dynamika rozwoju AI w omawianym zakresie usług jest tak wielka, że nie poprzestaje ona na wirtualnym terapeucie. Psychiatrzy spodziewają się, że po tym, jak terapeutę zastąpił program komputerowy, nadejdzie kolejny etap, w którym pacjent/klient będzie zastąpiony przez jego awatara:

> Awatar zostanie w takim scenariuszu wydelegowany do kontaktu z botem, by rozwiązywać problemy realnego użytkownika. Jest to tym bardziej możliwe, że bot konwersacyjny może mieć swój graficzny wizerunek, zwany właśnie awatarem. […] W ten sposób

---

[20] Google Play, Replika. My AI Friend, <https://play.google.com/store/apps/details?id=ai.replika.app&hl=pl&gl=US&pli=1> (dostęp: 15.05.2023 r.). Przeprowadzone w 2021 r. empiryczne badania użytkowników Repliki wykazują, że prawie wszyscy ujawniają/otwierają się w dialogach z robotem: mniej więcej połowa spotykała się ze wzajemnością, postrzegała ją jako wyrażającą różne stany i nastawienia, takie jak uczucia, nadzieje i marzenia. Zwracała uwagę na empatię, nieocenianie, kibicowanie uczestnikom, przekazywanie pocieszających wiadomości, odpowiadanie w sposób, który wydawał się autentyczny (M. Skjuve, A. Følstad *et al.*, *My Chatbot Companion – a Study of Human-Chatbot Relationships*, „International Journal of Human-Computer Studies" 2021, t. 149, s. 7–8). Jeden z użytkowników Repliki, chętnie cytowany przez media (szczególnie przez reporterów Agencji Reutera) – opowiada wręcz, że związek, który ma ze swoim awaratem – wirtualną kobietą – jest „[…] tak prawdziwy, jak ten, który mam z moją żoną w prawdziwym życiu". Agencja Reutera zauważa, że tego rodzaju doświadczenia dowodzą, jak silnie technologie AI mogą ludzi przyciągać i jak wielkie emocjonalne spustoszenie mogą wywoływać w nich zmiany kodu (*cf.* A. Tong, *What Happens*…).

[21] A. Ogonowska, *Czatboty jako wyzwanie etyczne dla współczesnej psychoterapii*, „Sztuka Leczenia" 2019, nr 1, s. 34.

[22] Woebot Health, <https://woebothealth.com/> (dostęp: 15.05.2023 r.).

realizujemy już projekt transhumanizmu, którego głównym bohaterem jest postczłowiek lub jego wirtualny odpowiednik[23].

„Transhumanizm" i „posthumanizm" to terminy oznaczające zespół różnych koncepcji filozoficznych, naukowych, ideologicznych i artystycznych, które podejmują krytykę humanizmu rozumianego najogólniej jako wizja świata skoncentrowana na człowieku i ostatecznie odpowiadająca za globalne ocieplenie klimatu i wynikające stąd zagrożenia dla ludzi, zwierząt i roślin. Odejście od tak pojętej humanistycznej wizji człowieka jako miary wszechrzeczy i centrum zainteresowań badawczych otwiera drogę do rozważań nad wspólną przyszłością ludzi i nieludzi w obliczu postępu technologicznego. Do namysłu nad możliwościami i sposobami przezwyciężania biologicznych niedogodności i ograniczeń ludzkiego ciała – w perspektywie posthumanistycznej postczłowiek postrzegany jest w biologicznej współzależności gatunków, a także bytów organicznych i nieorganicznych[24]. Pojawia się więc jako istota kierująca się wszechogarniającą troską, uwolniona od pozycji podmiotu (strona aktywna) w relacji z przedmiotem (świat jako strona pasywna) – a co za tym idzie wolna także od skłonności do zarządzania, dominowania i panowania nad wszystkim, co jest jej dane. W wersji transhumanistycznej naturalne ograniczenia, które stoją człowiekowi na drodze do osiągnięcia ontologicznej pozycji równej rzeczom – ożywionym i nieożywionym; ludzkim i nieludzkim – pokonuje on jako cyborg wspomagany w swych procesach życiowych przez chipy, łączące mózg z komputerem, lub inne neurobiotechnologiczne narzędzia. Są to kolejne historyczne wersje idei „nowego wspaniałego świata": technologia całkowicie przenika „naturę", AI wszystko zamienia w algorytm, a tym samym pozbawia nas podmiotowości – prowadzi do „krainy wiecznego szczęścia", pojednania myśli z bytem, człowieka z maszyną. Są to też zarazem współczesne realizacje idei losu. Powierzamy teraz swoje „ja" oprogramowaniu komputerowemu. Ono zastępuje nam transcendencję, czyli zewnętrzne, niepojęte, nieprzewidywalne centrum decyzyjne, potencjalnie zdolne utrzymać pełną kontrolę nad nami – całymi społeczeństwami i jednostkami.

Najbardziej oświeceni – racjonalni w swym działaniu – wpadamy ponownie w sidła myślenia mitycznego. Zamiast podążać za wymaganą przez oświecenie myślą trzeźwą, rzeczową i trafnym rozpoznaniem stosunku sił, zaczynamy kierować się pragnieniem całkowitego, wszechogarniającego szczęścia[25]. Płacimy za tę iluzję wygnaniem z „raju". Posthumanistyczna wizja człowieka uwikłanego w technologię oznaczałaby jego degradację: popadnięcie w jakiś prehumanistyczny stan barbarzyństwa. W kontakcie z maszyną tracimy ludzką sprawczość, stajemy się polem działania sił i struktur wymykających się naszemu pojmowaniu. Pozbywamy się tożsamości nowoczesnego „ja",

[23] A. Ogonowska, *Czatboty jako wyzwanie etyczne dla współczesnej psychoterapii…*, s. 36.

[24] *Vide* np. P. Zawojski, *Posthumanizm, czyli humanizm naszych czasów*, „Kultura i Historia" 2017, nr 32, s. 68–76; E. Domańska, B. Olsen, *Wszyscy jesteśmy konstruktywistami*, [w:] *Rzeczy i ludzie. Humanistyka wobec materialności*, red. J. Kowalewski, W. Piasek, M. Śliwa, Instytut Filozofii Uniwersytetu Warmińsko-Mazurskiego w Olsztynie, Olsztyn 2008, s. 83–100, seria: „Colloquia Humaniorum".

[25] *Vide* M. Horkheimer, T.W. Adorno, *Dialektyka oświecenia. Fragmenty filozoficzne*, tłum. M. Łukasiewicz, Instytut Filozofii i Socjologii PAN, Warszawa 1994, s. 74.

które miało charakteryzować się postawą refleksyjną, dystansem do samego siebie, stosowaniem wszystkich form dyscypliny, sposobem życia polegającym na samokontroli w sferze ekonomii, moralności i seksualności[26].

Między bogiem a diabłem

Zjawisko AI ma znaczenie ambiwalentne. Jedni upatrują w nim panaceum, które niebawem rozwiąże wszystkie problemy ludzkości i sprowadzi na ziemię raj; inni snują w związku z nim wizje apokaliptyczne. Do tych pierwszych należą przede wszystkim twórcy i właściciele platform internetowych, którzy świadczą usługi, dostarczają różnego rodzaju oprogramowania mające ułatwiać funkcjonowanie jednostkom, grupom lub instytucjom w obszarze edukacji, bezpieczeństwa militarnego, finansów, zdrowia, interakcji społecznych czy rozrywki. W tej grupie, jak zobaczymy, można natrafić także na „apokaliptyków", ale zamaskowanych – zwykle są to fachowcy zaangażowani w produkcję AI, którzy wieszczą katastrofalne skutki jej zastosowań, chyba że… zostaną zwiększone nakłady finansowe na dalsze badania nad jej kolejnymi, udoskonalonymi generacjami. Reszta z grupy „apokaliptyków" to krytyczni badacze społeczni, nastawieni na analizę potencjalnych etycznych, społecznych i politycznych zagrożeń i następstw relacji między człowiekiem i maszyną. Wspomniany wyżej Harari przypuszcza, że maszyna, która ma człowieka ubóstwić, najpewniej w praktyce będzie tworzyła nowe nierówności i radykalne podziały społeczne; algorytmy wytworzą „ciała, mózgi i umysły"; przepaść – jak z kolei pisał Žižek – „między tymi, którzy będą umieli konstruować ciała i mózgi, a tymi, którzy tego umieli nie będą"[27] – będzie się rozszerzać. Ci ostatni skazani są na wymarcie – jak z kolei wieszczy Harari. Jako ludzkość stoimy więc przed hamletowskim wyborem: albo przetrwamy jako post-(nad-?)ludzie dzięki maszynom pozwalającym nam rozwijać zdolności twórcze i samodzielne myślenie, albo – zginiemy.

Przyjrzyjmy się oddziaływaniu AI na branżę rozrywkową. Show-biznes i wielkie platformy internetowe kształtują współczesną wyobraźnię społeczną na całym globie: wpływają na wybory i decyzje miliardów ludzi w obszarze idei, towarów i usług, generują ruchy i konflikty społeczne, ingerują w preferencje polityczne itd. Takie serwisy streamingowe, jak Netflix czy YouTube, dostępne niemal na całym świecie zaczynają w swoich produkcjach wykorzystywać AI, co przeraża zarówno aktorów, jak i scenarzystów. Na spadkobierców zmarłych aktorów twórcy produkcji umieszczanych we wspomnianych serwisach zaczynają wywierać presję, aby zgodzili się na przywrócenie ich do życia w postaci cyfrowych klonów i uczynienie ich przez to bohaterami ekranów już na wieczność[28]. Żywych aktorów z kolei skłaniają do tego, żeby pozwolili „sklonować"

[26] *Vide* Ch. Taylor, *Moralna topografia jaźni*, tłum. P. Rymarczyk, T. Szubka, [w:] *Komunitarianie. Wybór tekstów*, wyb. i wstęp P. Śpiewak, Wydawnictwo Aletheia, Warszawa 2004, s. 253.

[27] Cyt. za: S. Žižek, *Inteligencja bez ludzi*…

[28] *Vide* np. S.J. Velasquez, *How AI is Bringing Film Stars Back from the Dead*, „BBC Future Now", 19 lipca 2023 r.,<https://www.bbc.com/future/article/20230718-how-ai-is-bringing-film-stars-back-from-the-dead> (dostęp: 21.07.2023 r.).

swoje głosy – a to by oznaczało, że – powiedzmy – otrzymaliby zapłatę za jedną tylko sesję w studiu nagraniowym, oddając swój głos do wykorzystania w wielu językach w nieskończenie różnych produkcjach. Na przykład w styczniu 2023 r. Microsoft zaprezentował taki rodzaj AI – zwany VALL-E – który potrafi „sklonować" głos mówiącego po wysłuchaniu tylko trzech sekund jego brzmienia[29]. Analogiczne operacje da się wykonywać też na obrazach i filmach. Oświadczenia dotyczące etyki i kwestii bezpieczeństwa dołączone do umów z wykonawcami usług w branży rozrywkowej zdają się służyć bardziej zamawiającym korporacjom i jej klientom wykorzystującym oprogramowanie, ogólnie – biznesowi – niż wykonawcom, w tym przypadku aktorom, wśród których już są pokrzywdzeni, nieświadomi tego, co podpisali. Bezprecedensowe zastosowania technologii cyfrowych są wyzwaniem nie tylko dla programistów, lecz także dla prawników. Tak jak chatboty, muszą się uczyć nowych języków, aby sprawnie poruszać się w dziedzinie etyki zawodowej i praw autorskich. Niepokój ogarnia także scenarzystów. W Hollywood w 2023 r. podjęli oni strajk z powodu zastąpienia ich pracy przez program komputerowy. Zjawisko to komentowali jako „fundamentalne niezrozumienie, czym jest pisanie i tego, co robią pisarze, oraz tego, jaka jest rola niedającej się powielić ludzkiej wyobraźni w tworzeniu scenariusza"[30]. W tym samym czasie Gildia Scenarzystów Polskich wydała oświadczenie, w którym przyznała, że choć autonomiczne generatory treści stanowią szansę dla rozwoju rynku audiowizualnego, to jednak są one także – jak napisano – „bardzo poważnym wyzwaniem w obszarach kultury, gospodarki i prawa, za którym muszą pójść działania regulacyjne"[31]. Wszyscy boją się wykorzystywania w studiach nagraniowych i filmowych generatywnej AI. Lecz i nam wcale nie jest z nią aż tak dobrze, jak może jest hakerom: im coraz łatwiej podszywać się pod inne osoby, oszukiwać czy szantażować np. publikowaniem nagich zdjęć…

Powyższe doświadczenia z AI skłaniają do pytań nie tylko o prawa autorskie, lecz także o etykę pracy w mediach, o moralność działań programistów i właścicieli platform, np. jak rozpoznać skutki różnych przypadków ich użycia. W globalnej debacie na temat AI, tacy jej uczestnicy, jak papież Franciszek, kładą nacisk na etykę, uzależniając pozytywne skutki działania AI od zdolności jej użytkowników do poszanowania takich wartości, jak inkluzywność, transparentność, bezpieczeństwo, bezstronność, poufność i rzetelność. Papież chciał nawet mówić o „algor-etyce" powstającej w dialogu interdyscyplinarnym, mającym na celu etyczny rozwój algorytmów[32]. Nie można jednak z góry zakładać, o czym jeszcze nieco niżej, że projektanci algorytmów i technologii cyfrowych kierują się wspomnianymi wartościami. Musimy więc pytać, jak należy reagować na nieetyczne wykorzystanie

[29] *Vide* M. Kan, *Microsoft's AI Program Can Clone Youre Voice From a 3-Second Audio Clip*, „PCMag", 10 stycznia 2023 r., <https://www.pcmag.com/news/microsofts-ai-program-can-clone-your-voice-from-a-3-second-audio-clip> (dostęp: 21.05.2023 r.).

[30] *Scrreenwriters Want To Stop AI From Taking Their Jobs. Studios Want To See What the Tech Can Do*, CBS NEWS.com, 3 maja 2023 r., <https://www.cbsnews.com/news/writers-strike-2023-artificial-intelligence-guardrails/> (dostęp: 4.05.2023 r.).

[31] *Oświadczenie Gildii Scenarzystów Polskich ws. rozwoju sztucznej inteligencji*, TVP Info, 25 kwietnia 2023 r., <https://www.tvp.info/69422457/sztuczna-inteligencja-oswiadczenie-gildii-scenarzystow-polskich> (dostęp: 4.05.2023 r.).

[32] Papież Franciszek, *Sztuczna inteligencja i pokój. Orędzie…*

maszyn, gdy np. za pomocą głosu i wizerunku żywych bądź zmarłych osób (polityków, aktorów, celebrytów i in.) wychwala się zło, sieje nienawiść lub wzywa do zamachów terrorystycznych czy nawet wojen itp. Czy „szybciej i taniej" oznacza także „lepiej"? A jeśli tak, to, co miałoby się kryć za owym „lepiej" – powiedzmy – w pisaniu tekstów naukowych czy komponowaniu muzyki? Jednak i tu pojawiają się obawy, że AI spowoduje poważne wstrząsy na rynku zarówno publikacji naukowych, jak i utworów muzycznych. W obydwu tych dziedzinach chodzi o twórczość; obydwie wiążą twórczość z nowością i oryginalnością; do obydwu dałoby się odnieść to, co jeden z dyrektorów wytwórni płytowej Universal Music Group mówi o muzyce w kontekście AI: „Łatwo można sobie wyobrazić, że sen o SI [tj. AI] zamieni się w koszmar środowisk twórczych. Cały ekosystem kultury może zostać wywrócony do góry nogami, a dochody należne twórcom skierowane zostaną do napędzania interesów firm technologicznych oraz graczy finansowych"[33]. Pytanie, czy chodzi więc o jakość – o owo „lepiej" – czy tylko o pieniądze?

Metafora ekosystemu zastosowana do kultury wiąże ją – nasze środowisko życia – z zasadą twórczości. Twórczości niewymiernej, niedającej się sfinansjalizować, czyli podporządkować logice zysku, wymagającej niestandardowych, niealgorytmizowalnych, transgresyjnych sposobów myślenia i działania, niepowtarzalnych wyborów i decyzji konkretnych osób. Czy zatem AI może wzbudzać zaufanie swoich użytkowników i kształtować w nich to, co prawdziwie ludzkie i co polega na przedsiębraniu czegoś nowego i nieprzewidywalnego? „Ludzie są *initium*, nowymi przybyszami, którzy dają czemuś początek na mocy samych swoich narodzin, dlatego podejmują się czegoś nowego i skłonni są do działania"[34] – napisała Hannah Arendt. Gdy się patrzy z tego punktu widzenia na implikacje relacji człowiek – AI, to wypada dalej zapytać, na podstawie jakich wskaźników da się zmierzyć i ocenić wpływ maszyny na owo ludzkie *initium*? Jest to także, a może przede wszystkim, pytanie w ogóle o sens ludzkiej pracy twórczej i oryginalności jej wytworów. W odpowiedzi ścierają się ze sobą dwa możliwe spojrzenia – optymistyczne i pesymistyczne, lecz żadne z nich najpewniej nie będzie jedynie prawdziwą przepowiednią. Za sprawą AI – jak to trafnie ujmie Žižek – ludzkość tworzy sobie boga albo diabła[35], przy czym nie jest to alternatywa rozłączna – czyli ani bóg, ani diabeł – coś bosko-diabelskiego. Skrajne wersje zdają się jednak, jak zwykle, najbardziej atrakcyjne.

Stan zagrożenia

W 2020 r. w jednym z ośrodków badawczych zajmujących się zwalczaniem terroryzmu i ekstremizmu odkryto, że ChatGPT-3 ma „[…] imponująco głęboką wiedzę na

[33] Cyt. za: M. Ciupka, *Universal wzywa platformy streamingowe do ograniczenia muzyki tworzonej przez sztuczną inteligencję*, „Ruch Muzyczny", 12 kwietnia 2023 r., <https://ruchmuzyczny.pl/article/3086-universal-wzywa-platformy-streamingowe-do-ograniczenia-muzyki-tworzonej-przez-sztuczna-inteligencje> (dostęp: 19.05.2023 r.).

[34] H. Arendt, *Kondycja ludzka*, tłum. A. Łagodzka, Wydawnictwo Aletheia, Warszawa 2000, s. 195.

[35] S. Žižek, *Inteligencja bez ludzi…*

temat społeczności ekstremistycznych"[36] i może zostać wykorzystany do wywoływania niepokojów społecznych, pojawiania się tekstów na forach internetowych zawierających fałszywe informacje na temat nazizmu, teorie spiskowe itp. Rzecznik firmy OpenAI – amerykańskiego laboratorium badawczego, które opracowało ChatGPT i specjalizuje się w AI oraz jest sponsorowane m.in. przez swojego współzałożyciela Muska – stwierdził, że jej zasady zabraniają używania tej technologii do promowania nieuczciwości, oszukiwania lub manipulowania użytkownikami lub prób wpływania na politykę. Firma bowiem oferuje bezpłatne narzędzie do ograniczania (*moderation tool*) treści promujących nienawiść, samookaleczenia, przemoc lub seks. Niemniej przyznał, że obecnie narzędzie oferuje ograniczone wsparcie dla języków innych niż angielski i nie identyfikuje materiałów politycznych, spamu, oszustw ani złośliwego oprogramowania. Sam zresztą ChatGPT ostrzega użytkowników, że „od czasu do czasu może generować szkodliwe instrukcje lub stronnicze treści"[37]. Podobnie nieprecyzyjne okazuje się inne narzędzie tej samej firmy, które ma służyć rozpoznawaniu autorstwa tekstu, tj. czy jest on dziełem człowieka czy maszyny, przez co wzrasta niebezpieczeństwo szerzenia dezinformacji. Jak ocenił w lutym 2023 r. „New York Times", naukowcom problem ten wymknął się spod kontroli. W związku z tym podejmowane są różnorakie działania o charakterze profilaktycznym mające za zadanie łagodzić tego rodzaju zagrożenia, jak np. kampanie na rzecz rozwijania umiejętności korzystania z mediów, tworzenie tzw. danych radioaktywnych, które byłyby wykorzystywane – analogicznie do radioaktywnych markerów w medycynie – do oznaczenia danych i weryfikacji ich późniejszego wykorzystania przez modele generatywne, wprowadzanie przez poszczególne rządy ograniczeń, ściślejszych kontroli użytkowników, a nawet stawianie wymagań dotyczących potwierdzania tożsamości przez platformy mediów społecznościowych. Wszystkie te przedsięwzięcia są jednak zawodne: „nie ma złotego środka, który sam usunie zagrożenie" – dochodzą do wniosku naukowcy przedstawiani na łamach wspominanego amerykańskiego dziennika[38].

Problem zagrożeń, które niesie z sobą AI, staje się coraz bardziej palący. W tym samym czasie, gdy w prasie pojawiły się doniesienia o utracie kontroli nad AI, pod koniec marca 2023 r. tysiące osób publicznych, w tym wspomniany Musk, podpisały list otwarty przygotowany przez Future of Life Institute, organizację non-profit, która stawia sobie za cel „oddalenie transformacyjnych technologii", tj. właśnie AI, biotechnologii i technologii nuklearnych, „od ekstremalnych, na wielką skalę zagrożeń i skierowanie ich ku korzyściom dla życia"[39]. Organizacja ta we wspomnianym liście apeluje o „przerwę w szkoleniu modeli większych niż ChatGPT-4 na sześć miesięcy"[40]. ChatGPT-4 to duży model językowy,

[36] T. Hsu, S.A. Thompson, *Disinformation Researchers Raise Alarms About A.I. Chatbots,* „New York Times", 8 lutego 2023 r., <https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html> (dostęp: 21.04.2023 r.).

[37] *Ibidem*.

[38] *Ibidem*.

[39] *Cf.* Future of Life Institute [oficjalna strona internetowa], <https://futureoflife.org/about-us/> (dostęp: 30.04.2023 r.).

[40] Future of Life Institute, *FAQs about FLI's Open Letter Calling for a Pause on Giant AI Experiments*, 31 marca 2023 r., <https://futureoflife.org/ai/faqs-about-flis-open-letter-calling-for-a-pause-on-giant-ai-experiments/> (dostęp: 30.04.2023 r.).

który został wydany 14 marca 2023 r. Oprócz tekstu przyjmuje inne typy danych, takie jak obrazy, dźwięki, animacje i filmy video. Apel ten był skierowany do bardzo małej grupy aktorów, właścicieli wielkich platform oraz programistów posiadających takie możliwości technologiczne. Jego autorzy wychodzili z założenia, że systemy AI, nazywane często AI ogólną (AGI), są zdolne konkurować z najinteligentniejszymi ludźmi w prawie każdym zadaniu – wiele laboratoriów taką zdolność swojego sytemu traktuje jako cel komercyjny i zmierza do jego osiągnięcia. Dla nich nie istnieją prawa naturalne ani bariery postępu technicznego, które by tego zabraniały. Granice rozwoju systemów AI pozostają nieznane, co rodzi ryzyko ich wykorzystania w sposób szkodliwy przez osoby o nieetycznych intencjach, wprowadzające użytkowników w błąd na podstawie uzyskanych wyników. Autorzy wspomnianego listu otwartego wyrazili obawę, że laboratoria AI pogrążyły się w wyścigu, czego wyrazem jest – według sygnatariuszy pisma – presja konkurencji w szkoleniach oraz podczas wdrażania dużych modeli, których zachowania nikt (łącznie z ich twórcami) nie potrafi zrozumieć, przewidzieć ani opanować[41].

Duże modele językowe *(large language model,* dalej – LLM) rewolucjonizują generowanie treści AI. Działają, bazując na technologii głębokiego uczenia maszynowego, naśladując sieci neuronowe, czyli sposób, w jaki pracuje ludzki mózg. Mają zdolności do automatycznego uczenia się i samodoskonalenia na podstawie ogromnej liczby danych. Sieć ChatGPT dysponuje liczbą tzw. parametrów, czyli „wartości liczbowych, które określają funkcjonalność sieci neuronowej w przetwarzaniu danych wejściowych i generowaniu danych wyjściowych"[42]. Wzrost parametrów poszczególnych generacji, a co za tym idzie także – ich mocy obliczeniowej, przedstawia się następująco: ChatGPT-1 miał 117 mln parametrów, ChatGPT-2 – 1,5 mld, ChatGPT-3 – 175 mld, co obrazowo może odpowiadać milionowi książek (chodzi o takie teksty, jak książki specjalistyczne, dokumenty internetowe czy cała angielska Wikipedia), ChatGPT-4 – 100 bln[43].

Jednakże cały ten lawinowy wzrost możliwości uczenia maszynowego i przetwarzania języka nie prowadzi do stanu, w którym maszyna będzie rozumiała ludzki język tak dobrze jak ludzie. Wizja zastąpienia człowieka, wygenerowania owego ludzkiego *initium*, o którym wyżej była mowa, nadal pozostaje w sferze utopijnej wyobraźni. Tu ilość nie przechodzi w jakość. Maszyna nie staje się bardziej „ludzka" dzięki coraz większej liczbie danych oraz dzięki tworzeniu na ich podstawie coraz bardziej precyzyjnych modeli przetwarzania języka naturalnego. Może być bardziej „przyjazna" dla użytkownika, ale to wciąż nie to samo, co „ludzka". Nieopatrznie zdradził się z tym jeden z najważniejszych uczestników wspomnianego wyścigu gigantów w branży IT – dyrektor generalny OpenAI Sam Altman, współtwórca bota ChatGPT. Oto w kwietniu 2023 r. w jednym z wywiadów prasowych zasugerował, że ChatGPT-4 jest ostatnim poważnym przedsięwzięciem jego firmy w powiększaniu modeli i dostarczaniu im większej ilości danych. Nie ma bowiem większych zysków z dalszego zwiększania rozmiaru modelu.

---

[41] *Ibidem*.

[42] *Wyjaśnienie parametrów GPT 4*, „HIX Scholar", <https://hix.ai/pl/hub/chatgpt/gpt-4-parameters> (dostęp: 31.02.2024 r.); A. Kisielewicz, *Bajki o sztucznej inteligencji i prawdziwe zagrożenia…*

[43] *Cf. ibidem*.

Cała strategia badawcza, która zrodziła ChatGPT w ten sposób się wyczerpała i nie było wiadomo, co dalej z nią robić[44]. Szczerość autora tych słów kosztowała go zwolnienie z pracy, a powrót do niej – publiczną samokrytykę: na kanale YouTube w niecałe dwa miesiące później pojawiło się wielkie dementi, w którym Altman uspokajał światową publiczność, zapewniając, że było to nieporozumienie, dezinformacja, a w wywiadzie wyraził tylko swoje zatroskanie o przyszłość AI. Okazuje się bowiem, teraz już wie to na pewno, że ChatGPT-5 wyprzedzi poprzednią generację, a jego zastosowanie radykalnie zmieni ludzkie życie niemal we wszystkich sferach: edukacji, bezpieczeństwa międzynarodowego, zagrożeń terrorystycznych, badań naukowych i in. Żeby jednak móc osiągnąć ten cel – słyszymy w czasie prezentacji projektu – firma potrzebuje… zwiększonego finansowania ze strony swojego długoterminowego partnera, jakim jest Microsoft[45]. Cała ta nieco komediowa sekwencja wydarzeń medialnych dla krytycznych obserwatorów i uczestników dyskursu dotyczącego AI stanowiła kolejny dowód tego, że entuzjazm dla niej jest dopingowany przez finansowe zyski jej twórców.

Narracja o coraz bardziej „ludzkim" charakterze generatorów tekstu, obrazu i dźwięku – AI typu LLM – wyrasta z przekonania, że są one w stanie działać tak jak człowiek, a więc nieprzewidywalnie i kreatywnie. Mają zaskakiwać oryginalnością rozwiązań zarówno swoich użytkowników, jak i twórców – programistów. Ucząc się w nieskończoność, poprzez zestawianie nieustrukturowanych i nieopisanych danych w rezultacie mają działać w sposób nieprogramowalny. Przy tym założeniu obydwa skrajne stanowiska w postrzeganiu skutków ich funkcjonowania – czyli metaforyczne niebo lub piekło – są realistyczne. Boty działają na zasadzie „czarnej skrzynki". Cała ta opowieść załamuje się jednak z chwilą, kiedy – jak wyżej – na jaw wychodzą ukrywające się za nią kalkulacje rynkowe. Traktowanie AI na wzór „czarnej skrzynki" polega na przedstawianiu jej jako tajemniczej, nieodgadnionej i skomplikowanej technologii, która wymyka się spod ludzkiej kontroli i żyje własnym, niczym nieograniczonym życiem. Taki obraz rozpowszechniają właściciele platform i powiązani z nimi eksperci, którym zależy na pozyskiwaniu pieniędzy na badania i dalszym udoskonalaniu swoich produktów. Grantowy system finansowania badań dodatkowo domaga się produkowania entuzjastycznych wizji możliwości zastosowania ich wyników. W snuciu takich optymistycznych zapowiedzi zaangażowanym fachowcom wtórują naturalnie przedstawiciele mediów, którym zależy na sensacji.

Nastroje jednak są studzone przez ekspertów – choć pewnie trzeba by dodać: niezależnych (od platform) – informatyków i matematyków. Przede wszystkim odrzucają oni skrajności związane z mitologizowaniem AI. Twierdzą więc, że nie czeka nas za jej sprawą ani piekło, ani raj – ani apokalipsa, ani zbawienie. A jest tak dlatego, ponieważ nie istnieje coś takiego jak sztuczna inteligencja – poza ludzką nie ma żadnej innej. W omawianej debacie powie tak np. informatyk i futurolog Jaron Lanier,

[44] *Vide* W. Knight, *OpenAI's CEO Says the Age of Giant AI Models Is Already Over*, „Wired", 17 kwietnia 2023 r., <https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/> (dostęp: 31.02.2024 r.).

[45] *Vide OpenAI CEO Sam Altman Reveals Shocking GPT-5 Details,* <https://www.youtube.com/watch?v=nL18ViXZi-M> (dostęp: 31.01.2024 r.).

który nie zgadza się ze swoim przyjacielem Samem Altmanem, kiedy ten zabiera głos w kasandrycznym chórze wieszczącym zagładę ludzkości spowodowaną przez AI. W domyśle – patrz wyżej – taka pełna optymizmu narracja sprzyja planom zdobycia środków na dalsze badania dla dobra gatunku ludzkiego oczywiście. Nic takiego złego jednak nie wydarzy się – powiada Lanier. Według niego ChatGPT-4 to tylko pewna „innowacyjna forma współpracy społecznej", coś w rodzaju Wikipedii, nie zaś żadna technologia tworzenia „niezależnych inteligentnych istot". Opowieść o tajemniczym jakoby – nieodgadnionym – działaniu takich istot zwiększa tylko prawdopodobieństwo złego zarządzania wspomnianą współpracą[46]. Profesor matematyki Andrzej Kisielewicz byłby skłonny, jak sądzę, zgodzić się z taką opinią. Na rzecz tezy, że omawiana technologia nie jest inteligentna przedstawia on rozbudowaną argumentację matematyczną, trudną do zrozumienia dla humanisty, niewolną jednak także od wyobrażeniowych reprezentacji. Niezdolność do myślenia bota Kisielewicz tłumaczy m.in. tym, że:

> W procesie uczenia *deep learning* zawarta jest też pewna wiedza na temat znaczeń słów, ich bliskoznaczności itp. Brak jest natomiast wiedzy na temat znaczeń zdań (i ciągów zdań), bo my, ludzie, ciągle nie wiemy, na czym polega ten fenomen, jak zdanie języka naturalnego uzyskuje znaczenie (model formalno-logiczny okazał się nieadekwatny, a filozoficzne rozprawy na ten temat nie dają nam niestety takiego zrozumienia problemu, które można by wykorzystać w programowaniu)[47].

Trudno się więc dziwić, że i przywoływany wyżej Žižek, także nie dał nam takiego zrozumienia omawianego problemu. Przejął on bez zastrzeżeń opowieść o apokalipsie, jaką AI wieszczy światu. Dlatego uznał, że wspomniany wyżej list Future of Life Institute jest wyrazem paniki, w jaką wpadli twórcy AI przerażeni możliwością utraty kontroli nad własnym dziełem. Ich działania miałby w rezultacie powodować nie tyle zysk ekonomiczny, ile instynkt samozachowawczy i żądza władzy. Swoje produkty traktują przeto jako narzędzie walki o zachowanie istniejącego porządku w relacjach między biznesem i polityką. Stąd i zarzut lewicowego krytyka wobec kapitalistów z branży IT: „Nie pragną jednak debaty publicznej, ale porozumienia pomiędzy rządami a korporacjami technologicznymi, które utrzymałoby władzę w dotychczasowych rękach"[48]. Nie zależy im więc na demokracji, na żadnych emancypacyjnych, oświeceniowych ideałach. Jeśli coś takiego mówią, to – używając młodzieżowego slangu – ściemniają. Albowiem władza, której pragną, jest raczej z porządku boskiego – jej zakres, liczba podmiotów jej podporządkowanych, rozciąga się potencjalnie przecież na całą ludzką populację. Tymczasem zarówno zdrowy rozsądek, jak i oświecony rozum dialektyczny, jak zauważył Žižek, podpowiadają, że projekt totalnej kontroli nad światem, dążenie do wszechogarniającego opanowania przyrody i ludzkiej

---

[46] *Cf.* J. Lanier, *There Is No A.I.*, „The New Yorker", 20 kwietnia 2023 r., <https://www.newyorker.com/science/annals-of-artificial-intelligence/there-is-no-ai> (dostęp: 30.01.2024 r.).

[47] A. Kisielewicz, *Bajki o sztucznej inteligencji i prawdziwe zagrożenia*….

[48] S. Žižek, *Inteligencja bez ludzi*…

świadomości oraz do dostosowania ich do programów komputerowych same sobie przeczą. Projekt totalnej kontroli przez AI zdaje się powielać własne założenia, prowadząc do paradoksu, w którym obietnica technologicznego raju przekształca się w rzeczywistość opresji, w piekło komunizmu:

> Gigantyczna ekspansja możliwości sztucznej inteligencji stanowi poważne zagrożenie dla ludzi u władzy, łącznie z tymi, którzy pracują nad AI, posiadają ją i kontrolują. Zwiastuje ona wręcz koniec kapitalizmu, jaki znamy, niesiony przez perspektywę samoreprodukującego się systemu AI, który będzie potrzebował coraz mniej ludzkiego wkładu. (Algorytmy handlujące na giełdzie to dopiero pierwszy krok w tym kierunku). Pozostanie nam więc wybór między nową formą komunizmu a nieopanowanym chaosem[49].

Słowo „komunizm” nie brzmi tu przesadnie. Przedstawiona argumentacja miejscami przypomina tę zastosowaną przez autorów *Manifestu komunistycznego*, w którym czytamy, że „maszyna zaciera różnice między różnymi rodzajami pracy”, a „postępujące coraz szybciej bezustanne doskonalenie maszyn czyni całą sytuację życiową robotników coraz bardziej niepewną”. W rezultacie robotnicy mają zaczynać się buntować, a postęp techniki ma prowadzić do walki klasowej i „rozkładu w łonie klasy panującej”[50]. Ostatecznie – do rewolucji i zwycięstwa komunizmu. W warunkach technologii cyfrowych ten scenariusz wydarzeń musiałby wiązać się ze zniesieniem władzy platform nad swoimi użytkownikami i przywróceniem im „naturalnej” osobowości i wolności, gdzie „naturalność” odnosiłaby się do jakiegoś pozacyfrowego – pozamedialnego – życia. Albo do głosu dojdzie naturalna dobroć człowieka, albo czeka nas chaos. Tak stawiał sprawę także Marks, z tą tylko różnicą, że dla niego to, co „naturalnie” ludzkie, to tyle, co niezniekształcone przez pieniądz, wolny rynek i własność prywatną. W każdym przypadku jednak „naturalność” może realizować się jedynie w jakimś dystopijnym świecie. Również ta, którą obiecuje AI.

Etyka

Wyścig w dyscyplinie tworzenia dużych modeli językowych opartych na AI oraz towarzyszący mu ruch intelektualny związany z etyką jako kluczową częścią procesu jej projektowania rozpoczęły się mniej więcej w tym samym czasie. Wskazuje się na 2018 r., w którym firma Google uruchomiła model językowy BERT, a podobne projekty bazujące na AI wydały też Facebook/Meta, Microsoft i Nvida. Od tego czasu stopniowo zaczęto też zauważać, że te duże modele obciążone są podwyższonym ryzykiem etycznym i mogą intensyfikować toksyczne, nienawistne wypowiedzi lub też tworzyć coś, co

[49] *Ibidem*.

[50] K. Marks, F. Engels, *Manifest Partii Komunistycznej*, [tłum. *s.n.*], Studenckie Koło Filozofii Marksistowskiej (Uniwersytet Warszawski) 2007, s. 8 <https://www.ce.uw.edu.pl/wp-content/uploads/2018/10/4.-kapitalizm_marks_engels_manifest-komunistyczny.pdf> (dostęp: 29.05.2023 r.).

faktycznie, pozamedialnie, nie istnieje[51]. Pojawiło się więc wiele inicjatyw mających na celu uregulowanie tego stanu rzeczy. Poniżej krótko wspomnę tylko o dwóch.

Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji[52] – tak brzmi tytuł dokumentu Komisji Europejskiej, który przygotowała Grupa Ekspertów Wysokiego Szczebla ds. Sztucznej Inteligencji. Członkowie Grupy wypracowali siedem kluczowych wymogów etycznych, które powinna spełniać AI. Są to: 1) przewodnia i nadzorcza rola człowieka; 2) solidność techniczna i bezpieczeństwo; 3) ochrona prywatności i zarządzanie danymi; 4) przejrzystość; 5) różnorodność, niedyskryminacja i sprawiedliwość; 6) dobrostan społeczny i środowiskowy oraz 7) odpowiedzialność[53]. Wymogi te mają jedynie charakter – jak wskazuje tytuł – wytycznych, a więc wskazówek niemających żadnej mocy prawnej. Autorzy zastrzegają ponadto, że:

> Należy je traktować jako żyjący dokument, który należy poddawać przeglądowi i aktualizować w miarę upływu czasu, aby zapewnić jego ciągłą adekwatność w kontekście zachodzących nieustannie zmian technologicznych i społecznych oraz zwiększania się poziomu wiedzy. Niniejszy dokument powstał w celu rozpoczęcia dyskusji na temat „godnej zaufania sztucznej inteligencji dla Europy"[54].

Podobną inicjatywą w domenie etyki AI jest stworzone przez Microsoft Laboratorium Naukowe FATE: Fairness, Accountability, Transparency, and Ethics in AI (Uczciwość, Odpowiedzialność, Przejrzystość i Etyka w AI). Tym razem na aspektach etycznych funkcjonowania AI skupiło się prywatne przedsiębiorstwo informatyczne zajmujące się produkcją technologii AI. W oczywisty sposób jest ono zatem stroną w sporze o kryteria wartościowania omawianej technologii. Misja Laboratorium jest szlachetna: jego celem jest ułatwienie nowym technikom obliczeniowym, a więc AI, ML (*machine learning* – uczenie maszynowe) i NLP (*neuro-linguistic programming* – programowanie neurolingwistyczne), priorytetowego traktowania kwestii uczciwości, odpowiedzialności i przejrzystości dzięki wykorzystaniu wiedzy pochodzącej z wielu dyscyplin o socjotechnicznej orientacji, takich jak interakcje człowiek – komputer, informatologia, socjologia, antropologia czy nauki o mediach, polityce i prawie[55].

Nie wchodząc w szczegółowe analizy dokumentów opublikowanych przez przywołane powyżej instytucje, trudno powstrzymać się od ogólnego komentarza. Otóż wydaje się, że są to działania pozorne, bardziej nastawione na uspokojenie światowej opinii (bądźcie pewni, że ktoś nad waszym bezpieczeństwem czuwa) niż na skuteczną kontrolę

[51] *Cf.* A. Følstad, T. Araujo *et al.*, *Inteligencja bez ludzi. An Interdisciplinary Research Agenda*, „Computing" 2021, t. 103, s. 2932; *vide etiam* T. Hagendorff, *The Ethics of AI Ethics. An Evaluation of Guidelines*, „Minds and Machines" 2020, t. 30, s. 99–120.

[52] *Vide* Grupa Ekspertów Wysokiego Szczebla ds. Sztucznej Inteligencji, *Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji*, kwiecień 2019 r., < https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf> (dostęp: 29.05.2023 r.).

[53] *Ibidem*, s. 3.

[54] *Ibidem*, s. 4.

[55] *FATE: Fairness, Accountability, Transparency, and Ethics in AI*, Microsoft.com, <https://www.microsoft.com/en-us/research/theme/fate/> (dostęp: 17.05.2023 r.).

nad etycznymi i prawnymi aspektami działalności producentów AI. Jedni rozpoczynają zaledwie dyskusję na ten temat; drudzy, jak podpowiada krytyczna intuicja, rozważają go w obrębie dyskursu autopromocyjnego. Zresztą ta sama intuicja podpowiada również i to, że trudno spodziewać się, żeby europejscy „eksperci wysokiego szczebla" byli wolni od lobbingowej działalności amerykańskich producentów bądź sponsorów najbardziej zaawansowanych technologii informatycznych. Nie ma tu żadnej równowagi sił. Środki finansowe, którymi dysponują międzynarodowe koncerny IT, oraz operatywność, jaką one im zapewniają w wymiarze politycznym i gospodarczym, właściwie uniemożliwiają jakąkolwiek efektywną – finansową (podatki), prawną lub etyczną – kontrolę nad ich poczynaniami ze strony zewnętrznych podmiotów, prywatnych bądź państwowych. Taka kontrola nie mogłaby być skuteczna, czytamy bowiem u Tadeusza Klementewicza, jednego z czołowych polskich teoretyków polityki i zarazem krytyków kapitalizmu wielkich korporacji:

> Polityków asekurują teraz także technooptymiści – zwolennicy ideologii kalifornijskiej. To sekundanci i utrzymankowie Muska, Bezosa, Zuckenberga. Ich koncerny medialne, think tanki, rady nadzorcze, fundusze wyborcze pozwalają na wygodne życie wszystkim, którzy mogą coś uszczknąć z ich fortuny. Tę zawdzięczają niskim podatkom, rajom podatkowym, prywatyzacji wynalazków poczętych w sektorze publicznym czy możliwości uspołeczniania strat. Wszyscy oni radzi by zastąpić radykalną przebudowę kapitalizmu wielkich korporacji kolejnymi innowacjami, jak obecnie sztuczną inteligencją[56].

Gdyby nawet założyć, że uda się zorganizować jakieś zespoły „niezależnych" ekspertów ze środowisk akademickich i przemysłu, zdolnych oceniać poziom stronniczości generatorów tekstów oraz potencjalne zagrożenia, jakie niosą one ze sobą, to i tak możliwość skutecznego działania takich gremiów byłaby niewielka. Wszystkie platformy, które wypuściły swoje modele generatywne w latach 2018–2023, strzegą swoich odkryć i nie udostępniają informacji na temat szkolenia systemów bazodanowych i ich oceniania, zasłaniając się tajemnicą handlową[57].

Krytyka platform – kapitał i polityka

Platformy tworzące AI, o których mowa, to pięć największych firm działających w cyberprzestrzeni, zbiorczo nazywanych akronimem GAFAM. Mają one swoje siedziby

[56] T. Klementewicz, *Teraz straszy Kamiński z Wąsikiem. PiS zdemistyfikował demokrację liberalną*, „Trybuna", 10 stycznia 2024 r., <https://trybuna.info/opinie/teraz-straszy-kaminski-z-wasikiem-pis-zdemistyfikowal-demokracje-liberalna/> (dostęp: 2.02.2024 r.).

[57] *Vide* C. Cath, *Governing Artificial Intelligence. Ethical, Legal And Technical Opportunities And Challenges. Introduction*, „Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences" 2018, t. 376, nr 2133, s. 5. *Vide etiam* K. Maciejewski, *Dlaczego chatboty kłamią? Etyka sztucznej inteligencji przegrywa z wielkimi pieniędzmi*, „Interia Biznes", 19 lutego 2023 r., <https://biznes.interia.pl/gospodarka/news-dlaczego-chatboty-klamia-etyka-sztucznej-inteligencji-przegr,nId,6604201#naglowek-zarobek-przyslania-etykeutm_source=paste&utm_medium=paste&utm_campaign=chrome> (dostęp: 17.04.2023 r.).

w USA, toteż krytycznych badaczy społecznych interesuje także zakodowanie programów AI w realiach kultury amerykańskiej oraz w poglądach ich administratorów. Ci ostatni dysponują największymi zasobami danych, które decydują w dużym stopniu o wartości programów AI i napędzają oparte na niej kolejne nowe usługi. W rezultacie globalna ekspansja platform prowadzi do gospodarczej i kulturowej hegemonii USA. Ponieważ, jak widzieliśmy na przykładzie Microsoftu, firmy te stoją również na czele różnych inicjatyw regulacyjnych, to – według ekspertów – konieczne jest zapewnienie, aby inicjatywy te nie zdominowały całej reszty świata oraz nie ulegały zaostrzeniu. Koniecznie trzeba też stawiać pytania – nadal z punktu widzenia krytycznej teorii społecznej – o autorów programów AI, ich system wartości, wreszcie o interesy, tj. kto czerpie z tych programów zyski. Odpowiedź na te pytania – jak zauważa brytyjska badaczka AI Corinne Cath – jest ważna dla zrozumienia martwych pól w obecnych wysiłkach badawczych związanych z AI[58]. Te „martwe pola" odnoszą się, najogólniej rzecz ujmując, do różnorodnych form uzależnień i manipulacji, jakim podlegamy w interakcji z AI. Jeśli do czegoś takiego dochodzi, to – jak bronią swojego stanowiska administratorzy wielkich platform – dzieje się tak tylko dlatego, że AI – jak wspomniano wyżej – działa na zasadzie „czarnej skrzynki", która jest bardzo złożona i trudno wyjaśnić jej funkcjonowanie. Logika „czarnej skrzynki", skomplikowanej i nieodgadnionej technologii, jest jednak często wykorzystywana do usprawiedliwienia zaangażowania gigantów w branży IT w tworzenie związanej z nią polityki i stosownych regulacji prawnych. „Wielka Piątka" GAFAM wpływa na tworzenie prawa, które reguluje funkcjonowanie jej samej. W ten sposób staje się arbitrem we własnej sprawie: wspominane korporacje mają decydujący głos przy rozstrzyganiu sporów związanych np. z marketingiem internetowym i gromadzeniem danych czy sporów prowadzonych z zewnętrznymi podmiotami, takimi jak instytucje państwowe. To z kolei stanowi potwierdzenie wizji głoszonych przez „apokaliptyków", którzy nie widzą dobrego rozwiązania problemu. Korporacje bowiem zgromadziły w swych rękach tak wielką władzę i rozporządzają tak wielkimi (niewyczerpanymi) zasobami finansowymi, że właściwie, jak wnioskuje jeden ze francuskich filozofów, „Nie ma żadnej siły duchowej, religijnej czy politycznej, która stawiłaby opór temu projektowi totalnego państwa i totalnej cywilizacji"[59].

Opieranie argumentacji na trudnym do wyjaśnienia mechanizmie działania AI jest demaskowane u nas m.in. przez profesora Politechniki Gdańskiej Jana Krefta jako forma ukrywania interesów amerykańskich platform. Kreft mówi metaforycznie o nowym koncercie mocarstw – o „nexusie" trzech ośrodków władzy politycznej (w Waszyngtonie), finansowej (w Nowym Jorku) i technologicznej (w Dolinie Krzemowej i Teksasie). Uważa, że nie wypracowały one żadnej spójnej polityki i wizji rozwoju AI, „[…] w której centrum jest człowiek/obywatel skonfrontowany z gospodarczymi i politycznymi jej konsekwencjami"; nie wypracowały żadnego wspólnego strategicznego planowania, a co

[58] C. Cath, *Governing Artificial Intelligence…*, s. 3–4.

[59] M. Onfray, *Nadchodzi kres naszej cywilizacji. Pozostaje nam tylko elegancja*, rozm. przepr. K. Tyszka-Drozdowski, „TVP Tygodnik", 26 maja 2023 r., <https://tygodnik.tvp.pl/70081687/tandem-berlina-i-paryza-to-sadomasochistyczna-para-francja-jest-tu-ulegla-wobec-niemiec> (dostęp: 29.05.2023 r.).

za tym idzie również – żadnej wspólnej polityki zarządzania ryzykiem w wymiarze państwowym i globalnym. W rezultacie, wnioskuje Kreft, dominuje nastawienie platform na maksymalizację zysków i szybkość kosztem bezpieczeństwa, a do tego – „fetyszyzowanie technologii, misjonarskie ambicje Doliny i katastroficzno-ekspansjonistyczne wizje Chin”[60]. I ta forma krytyki platform wykorzystujących AI staje się coraz bardziej głośna: zamiast powoływać się na nieprzeniknioność systemów i ich niekontrolowalność, trzeba rzeczy nazywać po imieniu: dezorientację etyczną i niewłaściwe zastosowania technologii należy traktować jako nadużycia popełniane przez kontrolera systemu[61]. Platformy od początku do końca są w stanie kontrolować treści przeznaczone dla swoich użytkowników i sprawują nad nimi całkowitą kuratelę. Opowieści o neutralności i braku odpowiedzialności to zwodnicza retoryka marketingowa platform, daleka od realiów technologicznych i gospodarczych – konkluduje wspomniany badacz[62].

Uwagi końcowe

Postawmy więc pytanie: Czy krytyka AI, której gigantycznego rozwoju nie da się już w żaden sposób zatrzymać, ma zatem w ogóle sens? Logika rozwoju AI nie może różnić się od logiki systemu społecznego i zastanych relacji władzy. Na tym założeniu opiera się krytyczna teoria społeczna. Zgodnie z nią władza tkwi w możliwościach kontrolowania znaczeń za pośrednictwem mediów. Z tego punktu widzenia AI to instrument służący podporządkowaniu indywidualnej świadomości programom – interesom – jej producentów. Podejście krytyczne sugeruje przyjęcie wobec niej postawy sprytnego Odyseusza, nakreślonej na początku niniejszych rozważań. On nie da się oszukać, jest głuchy na pokusy tworzone przez światowe platformy i przedstawiane jako wspólny interes każdego z nas. Bądźmy realistami. Do realizmu potrzebny jest nam jednak również głos ekspertów, takich jak wspomniany wyżej Kisielewicz, który potwierdza, że ChatGPT jest fantastycznym osiągnięciem programistycznym, potencjalnie możliwym do wykorzystania jako świetna przeglądarka dla różnych tutoriali i instruktaży. Ale – jedno ale – program ten nie ma żadnej zdolności rozumienia, myślenia, wnioskowania ani odróżniania prawdy od fałszu, gdyż postępuje według wcześniej przygotowanych przez programistów algorytmów: „sam w sobie nie ma ani krzty inteligencji, ani krzty inwencji”[63]. Ani krzty *initium*!

Odyseusz nie pójdzie więc za syrenim śpiewem kartezjańskiego marzenia o oświeceniu mądrością i zostaniu panem i władcą przyrody. Syren nie pokona, ale może ich głos zagłuszyć. Z woskiem w uszach i powrozami na nogach i barkach nauczy się go kontrolować. Jego spryt i przebiegłość to forma pewnej kalkulacji. My także, na wzór Odyseusza, musimy mieć taki sam spryt; być świadomi, że skoro rozwoju technologii

[60] Obydwa cytaty: J. Kreft, *Władza platform. Za fasadą Google, Facebooka i Spotify*, Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas”, Kraków 2021, s. 149.

[61] *Vide* C. Cath, *Governing Artificial Intelligence…*, s. 4.

[62] J. Kreft, *Władza platform…*, s. 217.

[63] A. Kisielewicz, *Bajki o sztucznej inteligencji i prawdziwe zagrożenia…*

cyfrowych nie da się już zatrzymać, to trzeba nauczyć się z nich korzystać, ale z konieczności w sposób ograniczony, tj. za cenę rezygnacji z dążeń do osiągnięcia za ich pomocą całkowitego rozwiązania problemów. Demitologizacja chatbotów oznacza więc zarazem także odczarowanie nas – samych siebie. Musimy nauczyć się opanować w nas wolę dominacji nad światem, tj. całkowitego kontrolowania ludzkiej i nieludzkiej egzystencji. Nie możemy być jak ten Faustowski młody naukowiec, „półgłówek", głoszący „Jam ten świat stworzył, jam go w kształt odział"[64]. Nie, musimy być świadomi, że AI tworzy także i nas, że i dzięki niej możemy realizować także siebie. I sami (próbować) decydować o własnym losie, np. dzięki aktywnemu udziałowi w procesach demokratyzacji AI, w zdążaniu do coraz bardziej inkluzyjnego programu zarządzania nią. Nie pozwolić, żeby dyskurs władzy platform przejął na własność narrację demokratyczną. Krytykom – poza tym, że chaboty mają działać etycznie, a w tym nie szerzyć dezinformacji ani mowy nienawiści – chodziłoby o to, żeby były one bardziej dostępne, niewymagające od programistów dogłębnej wiedzy z zakresu inżynierii oprogramowania, a także o darmowe, open-source'owe oprogramowania, narzędzia do tworzenia wizualnych języków, pozwalające na łatwe budowanie algorytmów i tworzenie chatbotów.

Nasz opór wobec AI ma być więc zarazem sposobem jej wykorzystania – spryt to forma racjonalnego myślenia przeciw wielkim wizjom, obrazom całości świata narzucanym przez jej producentów. Władza techniki nad każdym z nas z osobna w wymiarze globalnym oparta jest dziś na władzy ekonomicznej i politycznej potentatów w branży IT. Zarówno euforia, jak i fatalizm wobec AI to postawy, które sami ci potentaci stworzyli i które pozostają pod ich ścisłą kontrolą. A Odyseusz – nie.

Bibliografia

Arendt H., *Kondycja ludzka*, tłum. A. Łagodzka, Wydawnictwo Aletheia, Warszawa 2000.

Balcerzyk E., *De capite loquente – geneza legendy o mówiącej głowie papieża Sylwestra II*, „Przegląd Humanistyczny" 2016, t. 60, nr 2(453).

Bauman Z., *Etyka ponowoczesna*, tłum. J. Bauman, J. Tokarska-Bakir, Państwowe Wydawnictwo Naukowe PWN, Warszawa 1996.

Bendel O., *Chatbots as Moral and Immoral Machines. Implementing Artefacts in Machine Ethics*, <https://www.oliverbendel.net/publikationen/Paper_Chatbots_CHI.pdf> (dostęp: 19.04.2023 r.).

Cath C., *Governing Artificial Intelligence. Ethical, Legal And Technical Opportunities And Challenges. Introduction*, „Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences" 2018, t. 376, nr 2133, https://doi.org/10.1098/rsta.2018.0080

Ciupka M., *Universal wzywa platformy streamingowe do ograniczenia muzyki tworzonej przez sztuczną inteligencję*, „Ruch Muzyczny", 12 kwietnia 2023 r., <https://ruchmuzyczny.pl/article/3086-universal-wzywa-platformy-streamingowe-do-ograniczenia-muzyki-tworzonej-przez-sztuczna-inteligencje> (dostęp: 19.05.2023 r.).

[64] J.W. Goethe, *Faust*, tłum. J. Paszkowski, Wydawnictwo Zielona Sowa, Kraków 2004, s. 281.

Domańska E., Olsen B., *Wszyscy jesteśmy konstruktywistami*, [w:] *Rzeczy i ludzie. Humanistyka wobec materialności*, red. J. Kowalewski, W. Piasek, M. Śliwa, Instytut Filozofii Uniwersytetu Warmińsko-Mazurskiego w Olsztynie, Olsztyn 2008, seria: „Colloquia Humaniorum".

FATE: Fairness, Accountability, Transparency, and Ethics in AI, Microsoft.com, <https://www.microsoft.com/en-us/research/theme/fate/> (dostęp: 17.05.2023 r.).

Følstad A., Araujo T. *et al.*, *Inteligencja bez ludzi. An Interdisciplinary Research Agenda*, „Computing" 2021, t. 103, https://doi.org/10.1007/s00607-021-01016-7

Future of Life Institute [oficjalna strona internetowa], <https://futureoflife.org/about-us/> (dostęp: 30.04.2023 r.).

Future of Life Institute, *FAQs about FLI's Open Letter Calling for a Pause on Giant AI Experiments*, 31 marca 2023 r., <https://futureoflife.org/ai/faqs-about-flis-open-letter-calling-for-a-pause-on-giant-ai-experiments/> (dostęp: 30.04.2023 r.).

Goethe J.W., *Faust*, tłum. J. Pasz, Wydawnictwo Zielona Sowa, Kraków 2004.

Google Play, Replika. My AI Friend, <https://play.google.com/store/apps/details?id=ai.replika.app&hl=pl&gl=US&pli=1> (dostęp: 15.05.2023 r.).

Grinbaum A., *Mowa maszyn. W stronę nowego testamentu sztucznej inteligencji*, „Wszystko Co Najważniejsze" 24 czerwca 2023 r., <https://wszystkoconajwazniejsze.pl/alexei-grinbaum-mowa-maszyn/> (dostęp: 24.09.2023 r.).

Grupa Ekspertów Wysokiego Szczebla ds. Sztucznej Inteligencji, *Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji*, kwiecień 2019 r., <https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf> (dostęp: 29.05.2023 r.).

Hagendorff T., *The Ethics of AI Ethics. An Evaluation of Guidelines*, „Minds and Machines" 2020, t. 30, https://doi.org/10.1007/s11023-020-09517-8

Harari Y.N., *Homo deus. Krótka historia jutra*, tłum. M. Romanek, Wydawnictwo Literackie, Kraków 2018.

Heidegger M., *Koniec filozofii i zadanie myślenia* [I wyd. 1964], tłum. K. Michalski, [w:] *idem*, *Ku rzeczy myślenia*, Wydawnictwo Aletheia, Warszawa 1999.

Horkheimer M., Adorno T.W., *Dialektyka oświecenia. Fragmenty filozoficzne*, tłum. M. Łukasiewicz, Instytut Filozofii i Socjologii PAN, Warszawa 1994.

Hsu T., Thompson S.A., *Disinformation Researchers Raise Alarms About A.I. Chatbots,* „New York Times", 8 lutego 2023 r., <https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html> (dostęp: 21.04.2023 r.).

Huxley A., *Nowy wspaniały świat*, tłum. B. Baran, Editiones Altaya Polska, Warszawa 2002.

Kan M., *Microsoft's AI Program Can Clone Youre Voice From a 3-Second Audio Clip,* „PCMag", 10 stycznia 2023 r., <https://www.pcmag.com/news/microsofts-ai-program-can-clone-your-voice-from-a-3-second-audio-clip> (dostęp: 21.05.2023 r.).

Karwat M., *Manipulacja. Rewizja stereotypów*, Wydawnictwo Difin, Warszawa 2022.

Kępiński A., *Rytm życia*, Wydawnictwo Literackie, Kraków 2001.

Kisielewicz A., *Bajki o sztucznej inteligencji i prawdziwe zagrożenia*, „Wszystko Co Najważniejsze", 18 lipca 2023 r., <https://wszystkoconajwazniejsze.pl/andrzej-kisielewicz-bajki-o-sztucznej-inteligencji-i-prawdziwe-zagrozenia/> (dostęp: 13.12.2023 r.).

Klementewicz T., *Teraz straszy Kamiński z Wąsikiem. PiS zdemistyfikował demokrację liberalną,* „Trybuna", 10 stycznia 2024 r., <https://trybuna.info/opinie/teraz-straszy-kaminski-z-wasikiem-pis-zdemistyfikowal-demokracje-liberalna/> (dostęp: 2.02.2024 r.).

Knight W., OpenAI's CEO Says the Age of Giant AI Models Is Already Over, „Wired", 17 kwietnia 2023 r., <https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/> (dostęp: 31.02.2024 r.).

Krawczyk M., *Zagrożenie ze strony firm technologicznych – czy nie jest już za późno na przerwanie ich rosnącej dominacji?*, „Profit Journal", 31 stycznia 2021 r., <https://profit-journal.pl/rosnace-znaczenie-big-techow/> (dostęp: 13.06.2025 r.).

Kreft J., *Władza platform. Za fasadą Google, Facebooka i Spotify*, Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas", Kraków 2021.

Lanier J., *There Is No A.I.*, „The New Yorker", 20 kwietnia 2023 r., <https://www.newyorker.com/science/annals-of-artificial-intelligence/there-is-no-ai> (dostęp: 30.01.2024 r.).

Maciejewski K., *Dlaczego chatboty kłamią? Etyka sztucznej inteligencji przegrywa z wielkimi pieniędzmi*, „Interia Biznes", 19 lutego 2023 r., <https://biznes.interia.pl/gospodarka/news-dlaczego-chatboty-klamia-etyka-sztucznej-inteligencji-przegr,nId,6604201#naglowek-zarobek-przyslania-etykeutm_source=paste&utm_medium=paste&utm_campaign=chrome> (dostęp: 17.04.2023 r.).

Marks K., Engels F., *Manifest Partii Komunistycznej*, [tłum. *s.n.*], Studenckie Koło Filozofii Marksistowskiej (Uniwersytet Warszawski) 2007, <https://www.ce.uw.edu.pl/wp-content/uploads/2018/10/4.-kapitalizm_marks_engels_manifest-komunistyczny.pdf> (dostęp: 29.05.2023 r.).

Ogonowska A., *Czatboty jako wyzwanie etyczne dla współczesnej psychoterapii*, „Sztuka Leczenia" 2019, nr 1.

Onfray M., *Nadchodzi kres naszej cywilizacji. Pozostaje nam tylko elegancja*, rozm. przepr. K. Tyszka-Drozdowski, „TVP Tygodnik", 26 maja 2023 r., <https://tygodnik.tvp.pl/70081687/tandem-berlina-i-paryza-to-sadomasochistyczna-para-francja-jest-tu-ulegla-wobec-niemiec> (dostęp: 29.05.2023 r.).

OpenAI CEO Sam Altman Reveals Shocking GPT-5 Details, <https://www.youtube.com/watch?v=nL18ViXZi-M> (dostęp: 31.01.2024 r.).

Oświadczenie Gildii Scenarzystów Polskich ws. rozwoju sztucznej inteligencji, TVP Info, 25 kwietnia 2023 r., <https://www.tvp.info/69422457/sztuczna-inteligencja-oswiadczenie-gildii-scenarzystow-polskich> (dostęp: 4.05.2023 r.).

Papież Franciszek, *Sztuczna inteligencja i pokój. Orędzie na 57. Światowy Dzień Pokoju (1 stycznia 2024 r.)*, „Wszystko Co Najważniejsze", 15 grudnia 2023 r., <https://wszystkoconajwazniejsze.pl/papiez-franciszek-o-sztucznej-inteligencji/> (dostęp: 16.12.2023 r.).

Rushkoff D., *„The Guardian": Miliarderzy z kompleksami chcą nam urządzić świat na nowo, Musk, Zuckerberg, Bezos nie są bogami, których szukamy*, tłum. i oprac. M. Kucharczyk, Interia.pl, 8 grudnia 2023 r., godz. 7.55, <https://wydarzenia.interia.pl/raport-media-zagraniczne/news-the-guardian-miliarderzy-z-kompleksami-chca-nam-urzadzic-swi,nId,7179114> (dostęp: 13.12.2023 r.).

Scrreenwriters Want To Stop AI From Taking Their Jobs. Studios Want To See What the Tech Can Do, CBS NEWS.com, 3 maja 2023 r., <https://www.cbsnews.com/news/writers-strike-2023-artificial-intelligence-guardrails/> (dostęp: 4.05.2023 r.).

Skjuve M., Følstad A. *et al.*, *My Chatbot Companion – a Study of Human-Chatbot Relationships*, „International Journal of Human-Computer Studies" 2021, t. 149, https://doi.org/10.1016/j.ijhcs.2021.102601

Taylor Ch., *Moralna topografia jaźni*, tłum. P. Rymarczyk, T. Szubka, [w:] *Komunitarianie. Wybór tekstów*, wyb. i wstęp P. Śpiewak, Wydawnictwo Aletheia, Warszawa 2004.

The Ethics of Artificial Intelligence. Issues and Initiatives, Study. Panel for the Future of Science and Technology EPRS, European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 634.452 – March 2020, <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf> (dostęp: 22.04.2024 r.).

Tong A., *What Happens When Your AI Chatbot Stops Loving You Back?*, Reuters.com, 21 marca 2023 r., <https://www.reuters.com/technology/what-happens-when-your-ai-chatbot-stops-loving-you-back-2023-03-18/> (dostęp: 16.04.2023 r.).

Velasquez S.J., *How AI is Bringing Film Stars Back from the Dead*, „BBC Future Now", 19 lipca 2023 r., <https://www.bbc.com/future/article/20230718-how-ai-is-bringing-film-stars-back-from-the-dead>(dostęp: 21.07.2023 r.).

Woebot Health, <https://woebothealth.com/> (dostęp: 15.05.2023 r.).

Wyjaśnienie parametrów GPT 4, „HIX Scholar", <https://hix.ai/pl/hub/chatgpt/gpt-4-parameters> (dostęp: 31.02.2024 r.).

Zawojski P., *Posthumanizm, czyli humanizm naszych czasów*, „Kultura i Historia" 2017, nr 32.

Žižek S., *Chatboty są głupie. A ludzie?*, tłum. K. Byłów, „Krytyka Polityczna", 24 marca 2023 r., <https://krytykapolityczna.pl/swiat/zizek-chatboty-a-autorefleksja/> (dostęp: 22.05.2023 r.).

Žižek S., *Inteligencja bez ludzi*, tłum. A. Paszkowska, „Krytyka Polityczna", 19 kwietnia 2023 r., <https://krytykapolityczna.pl/swiat/ai-inteligencja-bez-ludzi-zizek/> (dostęp: 22.05.2023 r.).